K-means Clustering

In [113]:
```
import pandas as pd
import numpy as np
```

In [114]:
```
from sklearn.preprocessing import Imputer, LabelEncoder
from sklearn import preprocessing
import sklearn.metrics as sm
from sklearn.cluster import KMeans
```

In [115]:
```
names = ['Company','Open','High','Low','Close','Volume']
data = pd.read_csv('stockpredictions.csv', names=names)
print(data.shape)
```

```
(450, 6)
```

In [116]:
```
data.sample()
```

Out[116]:

	Company	Open	High	Low	Close	Volume
145	A	26.39	26.4552	26.32	26.34	4819

In [117]:
```
encoder = LabelEncoder()
data['Company'] = encoder.fit_transform(data['Company'])
```

In [118]:
```
data.sample(2)
names
```

Out[118]:
```
['Company', 'Open', 'High', 'Low', 'Close', 'Volume']
```

In [119]:
```
feature = ['Open', 'High', 'Low', 'Close', 'Volume']
target = ['Company']
```

In [120]:
```
x, y = data[feature], data[target]
```

In [145]:
```
x.sample()
```

Out[145]:

	Open	High	Low	Close	Volume
28	167.81	169.37	166.8	169.22	148597

In [146]:
```
model=KMeans(n_clusters=3)
```

In [147]:
```
model.fit(x)
```

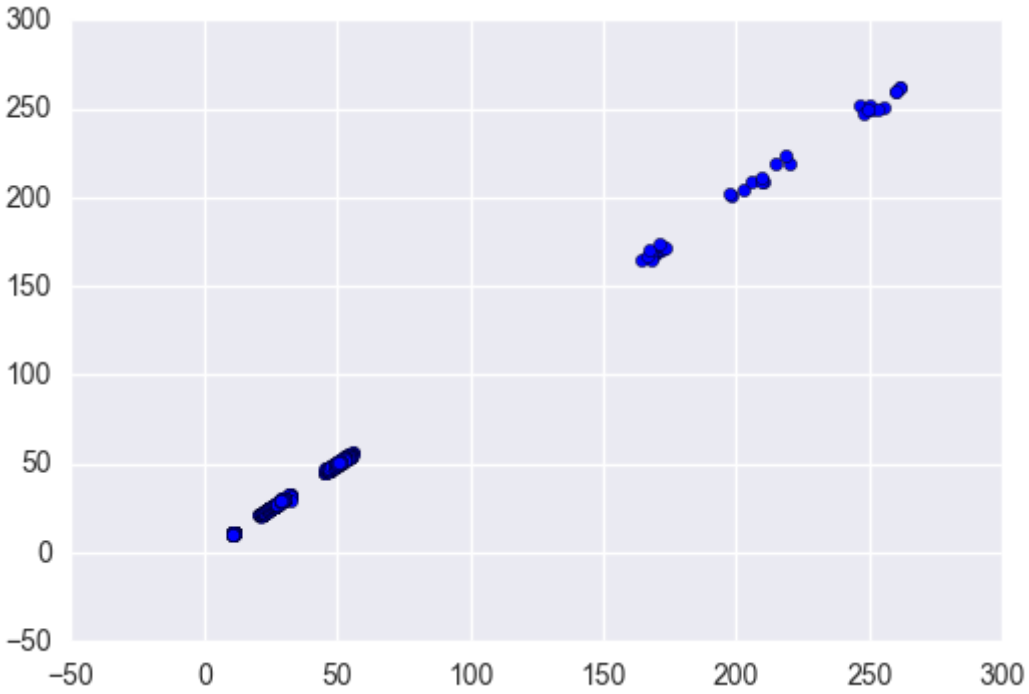
Out[147]:
```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
    n_clusters=3, n_init=10, n_jobs=1, precompute_distances='auto',
    random_state=None, tol=0.0001, verbose=0)
```

In [148]:
```
model.labels_
```

Out[148]:
```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 2, 2, 2, 1, 1, 1,
       2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 1, 2,
       1, 1, 2, 2, 1, 2, 1, 2, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 2, 2, 2, 1, 2, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       2, 2, 0, 0, 0, 0, 2, 0, 0, 0, 0, 2, 2, 0, 2, 0, 0, 0, 2, 0, 2, 0, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 2, 2, 2, 2, 2, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 0, 0, 0, 2, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 2, 2, 2, 0, 2, 2, 0, 2, 0, 0, 2,
       0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 2, 0, 0, 0, 2,
       2, 0, 0, 0, 2, 2, 2, 2, 0, 0, 0, 0, 2, 0, 0, 0, 0, 2, 0, 0, 2, 0, 2,
       0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 2, 2, 2, 2,
       2, 2, 0, 2, 2, 2, 0, 2, 2, 2, 2, 2, 0, 0, 2, 2, 0, 2, 0, 0, 0, 0, 2,
       0, 2, 2, 2, 0, 2, 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 2, 2, 2, 0, 0, 0, 2,
       0, 0, 2, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0])
```

In [149]:
```
plt.scatter(x.Open,x.Close)
```

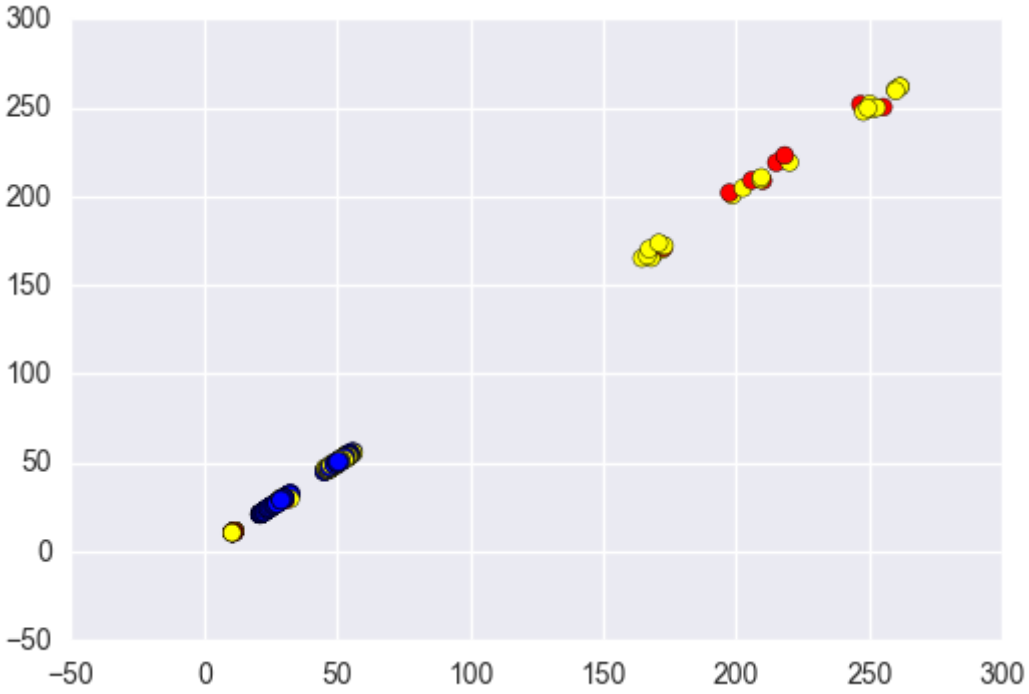
Out[149]: <matplotlib.collections.PathCollection at 0xe37d170>



In [150]:
```
color=np.array(['Blue', 'Red', 'Yellow'])
```

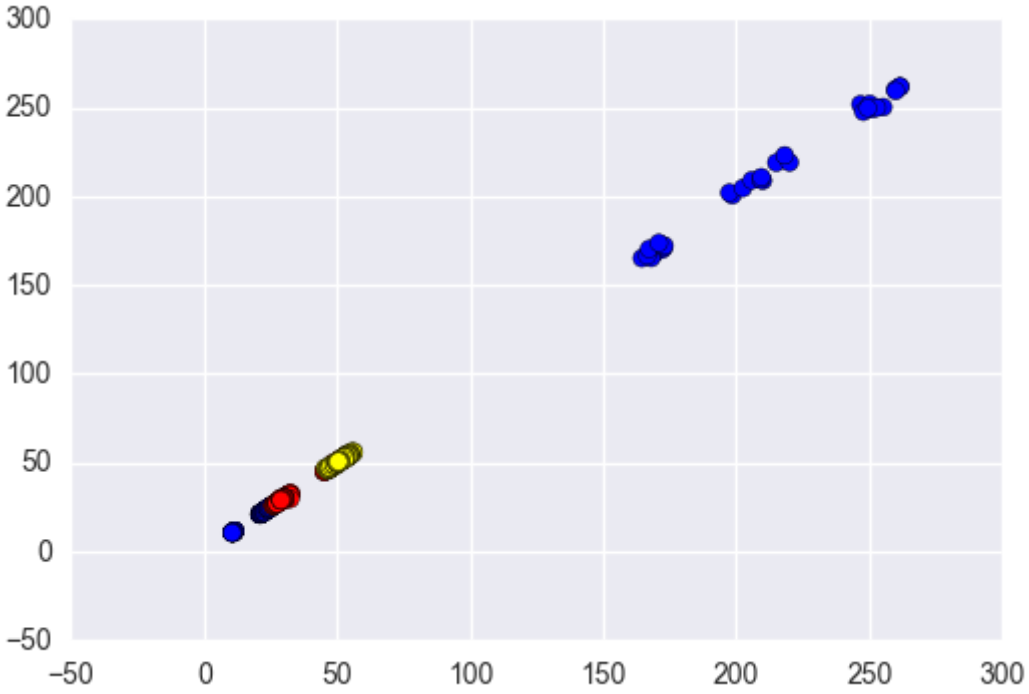
In [151]:
```
#Kmeans Clustering
plt.scatter(x.Open,x.Close,c=color[model.labels_],s=40)
```

Out[151]: <matplotlib.collections.PathCollection at 0xf350c70>



In [152]:
```
#Original Clustering
plt.scatter(x.Open,x.Close,c=color[y.Company],s=40)
```

Out[152]: <matplotlib.collections.PathCollection at 0xf3ebd30>



In [ ]: